

## How to use Variant Effects Report

- A. Introduction to Ensembl Variant Effect Predictor
  - B. Using [RefSeq\\_v1](#)
  - C. Using [TGACv1](#)
- 

### A. Introduction

The Ensembl Variant Effect Predictor is a toolset for the analysis, annotation, and prioritization of genomic variants in coding and non-coding regions. There are more than 1 million single nucleotide variants in wheat. There may be only 10 thousand that change the amino acid coding and a smaller subset of these that truncate or produce a loss of function.

The Ensembl Variant Effect Predictor (VEP) can be accessed by web, Perl script, web based API. The report page shows calculations for markers in T3 and provides links to calculation provided by Ensembl Plant

The inputs used for the VEP are

- a. Ensembl or VCF format SNP locations
- b. High confidence (HC) gene predictions
- c. FASTA file of the assembly

The outputs of the VEP are

- a. Feature – transcript (can also include motif and regulatory elements)
- b. Consequence
- c. Impact

A detailed description of the Consequence and Impact values can be found here:

[http://ensembl.org/info/genome/variation/predicted\\_data.html](http://ensembl.org/info/genome/variation/predicted_data.html)

The Ensembl VEP also incorporates SIFT and PolyPhen-2 but these are not used on the T3 website.

Sorting Intolerant From Tolerant ([SIFT](#)) predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physico-chemical similarity between the alternate amino acids.

SIFT value	Qualitative prediction	Website display example	
Less than 0.05	"Deleterious"	0.01	0.01
Greater than or equal to 0.05	"Tolerated"	0.8	0.8

[PolyPhen-2](#) predicts the effect of an amino acid substitution on the structure and function of a protein using sequence homology, Pfam annotations, 3D structures from PDB where available, and a number of other databases and tools (including DSSP, ncoils etc.)

## Consequence values for VEP

3\_prime\_UTR\_variant  
 5\_prime\_UTR\_variant  
 coding\_sequence\_variant  
 coding\_sequence\_variant,3\_prime\_UTR\_variant  
 coding\_sequence\_variant,5\_prime\_UTR\_variant  
 downstream\_gene\_variant  
 frameshift\_variant  
 frameshift\_variant,splice\_region\_variant  
 frameshift\_variant,splice\_region\_variant,intron\_variant  
 frameshift\_variant,start\_lost  
 frameshift\_variant,start\_lost,splice\_region\_variant  
 frameshift\_variant,start\_lost,start\_retained\_variant  
 frameshift\_variant,stop\_lost  
 frameshift\_variant,stop\_lost,splice\_region\_variant  
 frameshift\_variant,stop\_retained\_variant  
 inframe\_deletion  
 inframe\_deletion,splice\_region\_variant  
 inframe\_insertion  
 inframe\_insertion,splice\_region\_variant  
 inframe\_insertion,stop\_retained\_variant  
 intergenic\_variant  
 intron\_variant  
 protein\_altering\_variant  
 protein\_altering\_variant,splice\_region\_variant  
 splice\_acceptor\_variant  
 splice\_acceptor\_variant,5\_prime\_UTR\_variant  
 splice\_acceptor\_variant,coding\_sequence\_variant  
 splice\_acceptor\_variant,coding\_sequence\_variant,intron\_variant  
 splice\_acceptor\_variant,frameshift\_variant  
 splice\_acceptor\_variant,inframe\_insertion  
 splice\_acceptor\_variant,intron\_variant  
 splice\_donor\_variant  
 splice\_donor\_variant,coding\_sequence\_variant  
 splice\_donor\_variant,coding\_sequence\_variant,5\_prime\_UTR\_variant  
 splice\_donor\_variant,coding\_sequence\_variant,intron\_variant  
 splice\_donor\_variant,frameshift\_variant  
 splice\_donor\_variant,inframe\_insertion  
 splice\_donor\_variant,intron\_variant  
 splice\_region\_variant,3\_prime\_UTR\_variant  
 splice\_region\_variant,5\_prime\_UTR\_variant  
 splice\_region\_variant,intron\_variant  
 start\_lost  
 start\_lost,5\_prime\_UTR\_variant  
 start\_lost,inframe\_deletion  
 start\_retained\_variant  
 start\_retained\_variant,5\_prime\_UTR\_variant  
 stop\_gained  
 stop\_gained,frameshift\_variant  
 stop\_gained,frameshift\_variant,splice\_region\_variant  
 stop\_gained,inframe\_deletion  
 stop\_gained,inframe\_insertion  
 stop\_gained,inframe\_insertion,splice\_region\_variant  
 stop\_gained,splice\_region\_variant  
 stop\_lost,3\_prime\_UTR\_variant  
 stop\_lost,inframe\_deletion  
 stop\_retained\_variant,3\_prime\_UTR\_variant  
 upstream\_gene\_variant

### Impact values for VEP

HIGH - disruptive impact in the protein, protein truncation or loss of function

LOW – harmless, unlikely to change protein behaviour

MODERATE - non-disruptive variant that might change protein effectiveness

MODIFIER – usually non-coding variants

### General instructions for to use the T3 VEP report

First select a list of markers (limit the selection to under 1000). It will accept a mix of markers from different genotype experiments or a single genotype experiment. The positions of the markers on the genome assembly have been identified either by BLAST or from the coordinates provided when the genotype results were loaded into the database. If the marker position cannot be identified then it will be listed at the bottom of the page as not found.

### For markers not found on the map

You can run BLAST against RefSeq and format the output in either Ensembl or VCF format then mail the file to me using the feedback link on the T3 website. Then I can run the Ensembl VEP program on our machine and email you the results.

**B. Using RefSeq\_v1 assembly** – the only markers that have been mapped to RefSeq\_v1 are in the RefSeq v1.0 Physical Map and the 2017\_WheatCAP genotype experiment.

1. Visit <https://triticeaetoolbox.org/wheat/>
2. Select markers of interest. Go to Select => Markers

## Select Markers

### Currently selected markers

None

### Select markers by name

one or more markers

Synonyms will be translated.

Select by name

#### search using pattern matching

Synonyms will be translated.

. - matches any single character  
 \* - matches zero or more instances of preceding  
 ^ - matches at the beginning of value  
 \$ - matches at the end of value

Select by pattern matching

### Select markers in a range of map positions

Maps	Chromosome	Range	Markers
90K Array Consensus Aegilops tauschii, 2009 Chromosome Survey Sequence, 2014 CSS GBS 2014 CSS POPSEQ 2014 KleinProteo x KleinChaja, 2012 <b>RefSeq v1.0</b> SynOp GBS AntMap, 2012 SynOp GBS BinMap, 2012 TGACv1	RefSeq_1A RefSeq_1B RefSeq_1D RefSeq_2A RefSeq_2B RefSeq_2D RefSeq_3A RefSeq_3B RefSeq_3D RefSeq_4A	Map start: 198883 Map end: 830431481 Range: From 198883 to 830431481 Show markers	RAC875_c14696_364 RAC875_c14696_369 IWA3103 IWA3102 BS00084368_51 Excalibur_c2723_179 IWA5356 Kukri_c2972_110 BS00100045_51 Tdurum_contig14251_320 Tdurum_contig14251_421

Select markers

### 3. View the Variant Effects: Go to Reports => Variant Effects

#### Variant Effects

This page provides links to Sorting Intolerant From Tolerant (SIFT) and Variant Effect Predictor (VEP) to predict whether an amino acid substitution affects protein function. SIFT missense predictions for genomes: [Nature Protocols 2016; 11:1-9](#). The Ensembl Variant Effect Predictor: [Genome Biology Jun 6;17\(1\):122. \(2016\)](#) doi:10.1186/s13059-016-0974-4.

Genome Assembly

The links in the region column show known variations in a genome browser and their effects. The region is 1000 bases to either side of marker. The links in the gene column show a table with known variations, consequence type, and SIFT score.

marker	region	gene	description	feature	consequence	impact
<a href="#">RAC875_c11409_550</a>	<a href="#">chr3B:2958657</a>	<a href="#">TraesCS3B01G005000</a>	receptor kinase 1	-	intergenic_variant	MODIFIER
<a href="#">gbsCNLmaster28410</a>	<a href="#">chr3B:2959834</a>	<a href="#">TraesCS3B01G005000</a>	receptor kinase 1	-	intergenic_variant	MODIFIER
<a href="#">gbsHWWAMP34169</a>	<a href="#">chr3B:2959834</a>	<a href="#">TraesCS3B01G005000</a>	receptor kinase 1	-	intergenic_variant	MODIFIER
<a href="#">Excalibur_c3031_59</a>	<a href="#">chr3B:3036848</a>	<a href="#">TraesCS3B01G005100</a>	receptor kinase 1	TraesCS3B01G005100.1	downstream_gene_variant	MODIFIER
<a href="#">gbsHWWAMP55562</a>	<a href="#">chr3B:3042892</a>	<a href="#">TraesCS3B01G005100</a>	receptor kinase 1	TraesCS3B01G005100.1	upstream_gene_variant	MODIFIER
<a href="#">IACX4260</a>	<a href="#">chr3B:3228129</a>	<a href="#">TraesCS3B01G005800</a>	Paired amphipathic helix SIN3-like protein	TraesCS3B01G005800.1	frameshift_variant	HIGH
<a href="#">gbsHWWAMP49959</a>	<a href="#">chr3B:3228452</a>	<a href="#">TraesCS3B01G005800</a>	Paired amphipathic helix SIN3-like protein	TraesCS3B01G005800.1	frameshift_variant	HIGH

Selecting the link in the Gene column gives you a report of all markers for that gene.

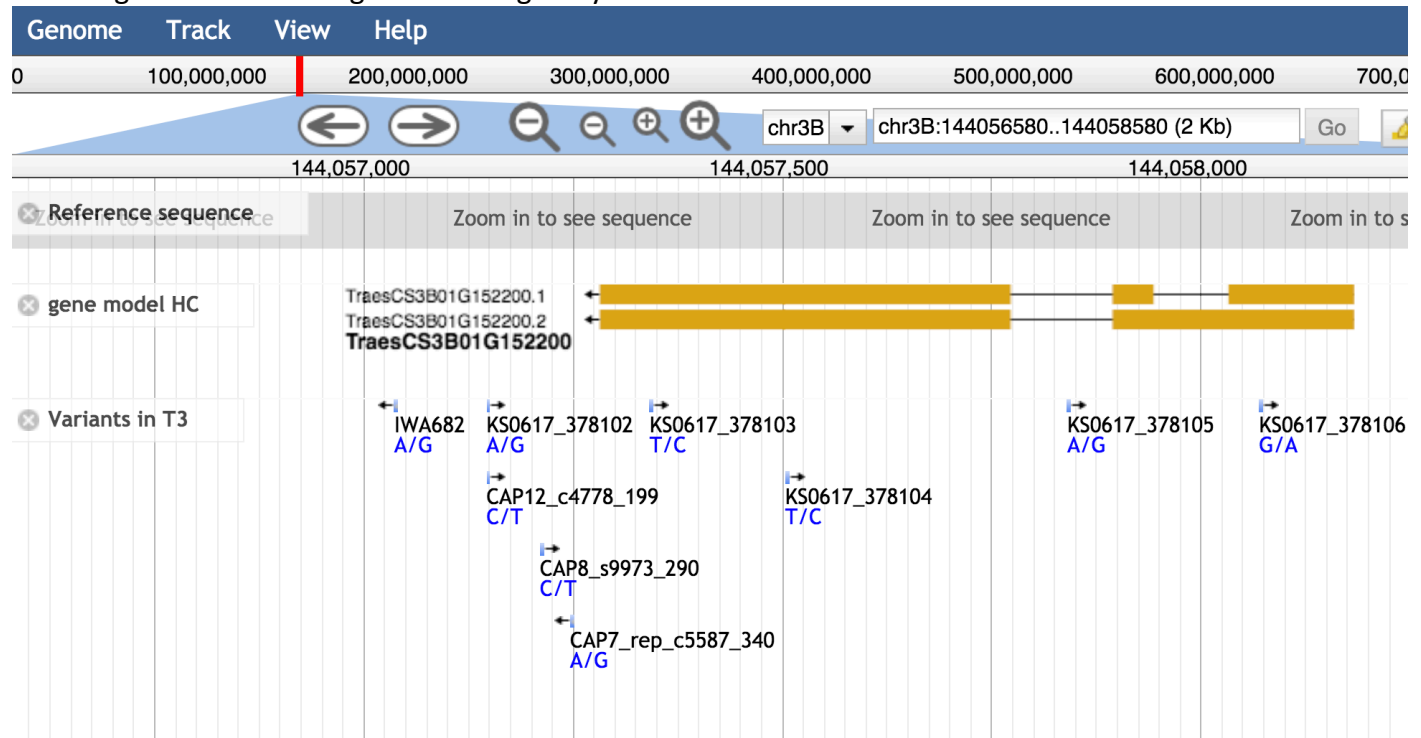
#### Variant Effects Gene TraesCS3B01G005800

ENSEMBL VARIANT EFFECT PREDICTOR v90.4

Citation: McLaren et. al. 2016 (doi:10.1186/s13059-016-0974-4)

marker name	location	feature	consequence	impact
<a href="#">gbsHWWAMP49959</a>	<a href="#">chr3B:3228452</a>	<a href="#">TraesCS3B01G005800.1</a>	frameshift_variant	HIGH
<a href="#">IACX4260</a>	<a href="#">chr3B:3228129</a>	<a href="#">TraesCS3B01G005800.1</a>	frameshift_variant	HIGH
<a href="#">KS0617_361198</a>	<a href="#">chr3B:3225869</a>	<a href="#">TraesCS3B01G005800.1</a>	5_prime_UTR_variant	MODIFIER
<a href="#">KS0617_361199</a>	<a href="#">chr3B:3225892</a>	<a href="#">TraesCS3B01G005800.1</a>	splice_region_variant,5_prime_UTR_variant	LOW
<a href="#">KS0617_361200</a>	<a href="#">chr3B:3225927</a>	<a href="#">TraesCS3B01G005800.1</a>	intron_variant	MODIFIER
<a href="#">KS0617_361201</a>	<a href="#">chr3B:3225929</a>	<a href="#">TraesCS3B01G005800.1</a>	intron_variant	MODIFIER
<a href="#">KS0617_361202</a>	<a href="#">chr3B:3225933</a>	<a href="#">TraesCS3B01G005800.1</a>	intron_variant	MODIFIER

Selecting the link in the region column gives you a JBrowse view for that marker



## C. Using TGACv1 assembly

1. Select markers of interest. Go to Select => Markers

### Select Markers

**Currently selected markers**

None

**Select markers by name**

one or more markers

Synonyms will be translated.

Select by name

**search using pattern matching**

Synonyms will be translated.

Select by pattern matching

. - matches any single character  
 \* - matches zero or more instances of preceding  
 ^ - matches at the beginning of value  
 \$ - matches at the end of value

**Select markers in a range of map positions**

Maps	Chromosome	Range	Markers
90K Array Consensus Aegilops tauschii, 2009 Chromosome Survey Sequence, 2014 CSS GBS 2014 CSS POPSEQ 2014 KleinProteo x KleinChaja, 2012 RefSeq v1.0 SynOp GBS AntMap, 2012 SynOp GBS BinMap, 2012 TGACv1	TGACv1_1A TGACv1_1B TGACv1_1D TGACv1_2A TGACv1_2B TGACv1_2D TGACv1_3A TGACv1_3B TGACv1_3D TGACv1_4A	Map start: 29 Map end: 562460 Range: From 29 to 562460 Show markers	synopGBS114094 BobWhite_c32624_354 Excalibur_rep_c70950_414 RAC875_c8081_165 Kukri_c4744_265 gbsHWWAMP2857 gbsCNLmaster14715 Kukri_c33749_590 Kukri_s110000_156 JD_c25161_410

Select markers

2. View the Variant Effects: Go to Reports => Variant Effects
3. Select TGACv1 for the Genome Assembly
4. To view the Variant Effect, you can either
  - a. Click on the link in the Gene column. This will take you to a table on Ensembl Plant website.
  - b. Scroll down the page until you see the second table. Copy these entries and past them in the data field of the Variant Effect Predictor on the Ensembl Plant website.

## Variant Effects

This page provides links to Sorting Intolerant From Tolerant (SIFT) and Variant Effect Predictor (VEP) to predict whether an amino acid substitution affects protein function. SIFT missense predictions for genomes: [Nature Protocols 2016; 11:1-9](#). The Ensembl Variant Effect Predictor: [Genome Biology Jun 6;17\(1\):122. \(2016\)](#) doi:10.1186/s13059-016-0974-4.

Genome Assembly  To access additional assemblies [Login](#).

Warning: no local VEP calculations for TGACv1, use the link in the gene column to show a table with known variations. The links in the region column show known variations in a genome browser and their effects. The region is 1000 bases to either side of marker. The links in the gene column show a table with known variations, consequence type, and SIFT score.

marker	region	gene	description
<a href="#">gbsCNLmaster43846</a>	2B:115	<a href="#">TRIAE_CS42_2BS_TGACv1_150331_AA0498120</a>	
<a href="#">gbsHWWAMP53077</a>	2B:115	<a href="#">TRIAE_CS42_2BS_TGACv1_150331_AA0498120</a>	
<a href="#">RAC875_c25282_67</a>	2B:119	<a href="#">TRIAE_CS42_2BS_TGACv1_150331_AA0498120</a>	
<a href="#">BobWhite_c37770_79</a>	2B:120	<a href="#">TRIAE_CS42_2BS_TGACv1_150331_AA0498120</a>	
<a href="#">gbsCNLmaster45422</a>	2B:90	<a href="#">TRIAE_CS42_2BS_TGACv1_150331_AA0498120</a>	
<a href="#">gbsHWWAMP55080</a>	2B:90	<a href="#">TRIAE_CS42_2BS_TGACv1_150331_AA0498120</a>	

To run Variant Effect Predictor, copy the data below and paste it into the text box on the website [Ensembl Plant VEP](#). Calculations take about 5 minutes per marker.

TGACv1\_scaffold\_150331\_2BS 90 90 A/G + [gbsCNLmaster45422](#)

TGACv1\_scaffold\_150331\_2BS 90 90 A/G + [gbsHWWAMP55080](#)

TGACv1\_scaffold\_149393\_2BS 115 115 C/G + [gbsCNLmaster43846](#)

TGACv1\_scaffold\_149393\_2BS 115 115 C/G + [gbsHWWAMP53077](#)

TGACv1\_scaffold\_152613\_2BS 119 119 A/G - [RAC875\\_c25282\\_67](#)

TGACv1\_scaffold\_144458\_2BL 120 120 T/C + [BobWhite\\_c37770\\_79](#)



Selecting the link in genes column directs you to Ensembl Plant to show you a table of variants for that gene.


Filter <span>▼ SIFT: All</span> <span>▼ Consequences: All</span> <span>▼ Filter Other Columns</span>							
Show/hide columns <span>Search...</span>							
Variant ID	Chr: bp	Alleles	Source	Conseq. Type	SIFT	Transcript	
<a href="#">Cadenza1269:3379728</a>	TGACv1_scaffold_24641 8_3B:20	G/A	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza0447:3379729</a>	TGACv1_scaffold_24641 8_3B:21	C/T	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza0779:3379730</a>	TGACv1_scaffold_24641 8_3B:23	G/A	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza1090:3379731</a>	TGACv1_scaffold_24641 8_3B:28	G/A	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza1174:3379732</a>	TGACv1_scaffold_24641 8_3B:35	G/A	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza2068:3379733</a>	TGACv1_scaffold_24641 8_3B:54	C/T	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza1598:3379734</a>	TGACv1_scaffold_24641 8_3B:62	C/T	EMS-induced mutation	5 prime UTR variant	-	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza1410:3379736</a>	TGACv1_scaffold_24641 8_3B:106	C/T	EMS-induced mutation	missense variant	0	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza0289:3379735</a>	TGACv1_scaffold_24641 8_3B:106	C/T	EMS-induced mutation	missense variant	0	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	
<a href="#">Cadenza1533:3379737</a>	TGACv1_scaffold_24641	C/T	EMS-induced	missense variant	0.02	<a href="#">TRIAE_CS42_3B_TGACv1_246418_AA0834350.1</a>	

On the Ensembl page there is a link to view the location in the Ensembl Browser, which will show you the position and variant types for that gene



You can run the Ensembl VEP program for your own markers by copying the results from the bottom table of the T3 Variant Effects page and pasting it into the tool on the Ensembl Plant website. [http://plants.ensembl.org/Triticum\\_aestivum/Tools/VEP?db=core](http://plants.ensembl.org/Triticum_aestivum/Tools/VEP?db=core)

### Variant Effect Predictor

**Species:** Triticum aestivum 

**Assembly:** TGACv1

**Name for this job (optional):**

**Either paste data:**

```

245995_3B 276 276 T/G - JD_c58431_362
732361_3B 284 284 T/C + RAC875_c53667_348
235132_3B 285 285 A/C - gbsCNLmaster7917
235132_3B 285 285 A/C - gbsHWWAMP9550
235132_3B 286 286 A/C - synopGBS105367
234983_3B 288 288 A/G + gbsHWWAMP26271


```


[Run instant VEP for current line >](#)


**Examples:** [Ensembl default](#), [VCF](#), [Variant identifiers](#), [HGVS notations](#)

**Or upload file:** Choose File No file chosen

**Or provide file URL:**

**Identifiers and frequency data**  *Additional identifiers for genes, transcripts and variants; frequency data*

**Extra options**  *e.g. SIFT, PolyPhen and regulatory data*

**Filtering options**  *Pre-filter results by frequency or consequence type*

Run > [Clear](#)